

39

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/00, 15/86, A01N 63/00	A1	(11) International Publication Number: WO 00/56874 (43) International Publication Date: 28 September 2000 (28.09.00)
(21) International Application Number: PCT/US00/07841 (22) International Filing Date: 24 March 2000 (24.03.00) (30) Priority Data: 60/126,123 25 March 1999 (25.03.99) US (71) Applicant: MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; Five Cambridge Center, Room NE 25-230, Kendall Square, Cambridge, MA 02142-1493 (US). (72) Inventors: HOPKINS, Nancy; 975 Memorial Drive, Cambridge, MA 02138 (US). CHEN, Wenbiao; 268 Grove Street, Newton, MA 02466 (US). BURGESS, Shawn; 372 Silver Street, Boston, MA 02110 (US). AMSTERDAM, Adam; 26 Hurlbut Street #5, Cambridge, MA 02138 (US). (74) Agent: BIEKER-BRADY, Kristina; Clark & Elbing, LLP, 176 Federal Street, Boston, MA 02110-2214 (US).		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: RETROVIRAL VECTORS AND METHODS FOR PRODUCTION AND USE THEREOF		
(57) Abstract <p>The invention features methods and reagents for introducing a mutation into a gene in a cell. The method includes contacting the cell with a recombinant retrovirus including: (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; and (v) viral long-terminal repeats, wherein the splice acceptor and the splice donor flank nucleic acid sequence encoding a stop codon that is in frame with the splice acceptor; and allowing the retrovirus to integrate into a gene of the cell. Integration of the retrovirus into the gene introduces a mutation into the gene.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

RETROVIRAL VECTORS AND METHODS
FOR PRODUCTION AND USE THEREOF

5 **Statement as to Federally Sponsored Research**

This research has been sponsored in part by NIH grants 5-RO1-CA19308-22 and 1-RO1-RR12589-01. The government has certain rights to the invention.

Background of the Invention

10 The invention relates to recombinant retroviral vectors, and the production and uses thereof.

A powerful approach to understanding the genetic basis of developmental processes is the application of classical Mendelian genetics. Although long considered practical only in invertebrate animals, recently, large
15 scale metagenesis screens to identify embryonic lethal genes have been completed in a vertebrate animal, the zebrafish. In these screens, mutations were induced by chemical mutagens, which cause single nucleotide changes in the animal's DNA. A mutagenesis approach was feasible in the fish since it is possible to breed and maintain very large numbers of fish in the lab, and
20 because early developmental mutations are easy to identify in fish embryos since these embryos develop outside the mother and are transparent for the first week of life.

The results from the mutagenesis screens suggest that there are approximately 2400 embryonic lethal genes in the fish. Loss-of-function
25 mutations in any of these genes result in embryonic lethality by about five days of age. Mutants fall into three classes: about 20% of all mutants display apoptosis in the CNS; about 50% of all mutants display defects in multiple organs or structures; and about 30% of all mutants have developmentally

specific mutations that affect primarily one or just a few organs. Mutations that affect the development of virtually every embryonic organ system that can be seen by low power microscopy have been found. Another large class of embryonic lethal mutations affect motility of the embryo.

5 Because the mutagen used in the aforementioned studies causes only single nucleotide changes, mutated genes must be cloned by positional cloning or by candidate gene cloning, both of which are costly and time-consuming. This problem is compounded by the fact that, while the zebrafish genome is large (approximately two-thirds the size of the mouse genome), the zebrafish
10 genome project is relatively poorly developed.

 The use of candidate genes relies on prior knowledge of the expression pattern or function of a gene and a correlation between these attributes and the mutant phenotype. This approach, by definition, is strongly biased against isolation of genes that have no highly-conserved orthologue.
15 Moreover, it is likely that one could screen tens of candidate gene without finding the mutated gene.

 To increase the ease of detecting mutations in genes, it would be desirable to develop a retroviral vector that carries a gene trap. A gene trap construct harbors a nucleic acid sequence, such as a reporter gene, that is
20 expressed only when the virus integrates into an active gene. The nucleic acid sequence contains a splice acceptor at the 5' end, allowing for transcription and translation of both the gene that received the insertion and the inserted nucleic acid sequence itself. Introduction of a mutation (a stop codon or a frameshift) after the reporter gene results in truncation (and possibly loss of function) of
25 the interrupted protein.

 Strategies for mutating genes using a retroviral gene trap vector have been attempted in vertebrates, particularly in mice. In spite of the many advantages of using retroviruses to introduce gene trap cassettes, there has been

very little success. Hence there is a need for improvement of the current methods and/or gene trap vectors.

It would also be advantageous to develop high-titer virus producer cell lines. Such cell lines are not only useful for mutagenesis of animals with retroviral vectors, but would also permit efficient production of retroviral
5 vectors for the construction of transgenic animals and for human gene therapy.

Summary of the Invention

In a first aspect, the invention features a recombinant retrovirus including: (a) branch-point sequence; (b) a polypyrimidine tract; (c) a splice
10 acceptor; (d) a splice donor; and (e) LTRs. Preferably, the splice acceptor and the splice donor flank nucleic acid sequence encoding a stop codon that is in frame with the splice acceptor.

In a preferred embodiment, the retrovirus includes a reporter gene such as *gfp*, *lacZ*, or a nucleic acid encoding myc epitope, a FLAG epitope, or a
15 HA epitope. The reporter gene is preferably in the direction opposite to the direction of transcription from the viral long-terminal repeats. In one preferred embodiment, there are reporter genes in all three reading frames.

In other preferred embodiments, the retrovirus includes a splice enhancer (e.g., a splice enhancer from the avian sarcoma leukosis virus) or
20 exonic sequence between the splice acceptor and the splice donor.

In still another preferred embodiment, the retrovirus includes nucleic acid sequence encoding a polypeptide encoded in the direction opposite to the direction of transcription from the viral long-terminal repeats. Exemplary polypeptides include, but are not limited to, GFP, β -galactosidase, a myc
25 epitope, a FLAG epitope, a HA epitope, Cre recombinase, and FLP recombinase.

In a second aspect, the invention features a method for performing gene-trapping in a cell. the method includes (a) contacting the cell with a recombinant retrovirus that includes (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-terminal repeats; and (vi) a reporter gene in an orientation opposite to the direction of transcription from the viral long-terminal repeats; and (b) allowing the retrovirus to integrate into the genome of the cell. The reporter gene is expressed if there is a gene-trapping event. The cell may be *in vitro* or *in vivo*.

In a third aspect, the invention features a method for introducing a mutation into a gene in a cell, including (a) contacting the cell with a recombinant retrovirus including: (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; and (v) viral long-terminal repeats, wherein the splice acceptor and the splice donor flank nucleic acid sequence encoding a stop codon that is in frame with the splice acceptor; and (b) allowing the retrovirus to integrate into a gene of the cell. Integration of the retrovirus into the gene introduces a mutation into the gene. The method may also include (c) determining the site of integration of the retrovirus. The cell may be *in vitro* or *in vivo*.

In a fourth aspect, the invention features a method for determining the expression pattern of a gene in a non-human animal, including (a) introducing into the animal or an ancestor thereof a recombinant retrovirus including (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-terminal repeats; and (vi) nucleic acid sequence between the splice acceptor and the splice donor, the nucleic acid sequence encoding a polypeptide in the direction opposite to the direction of transcription from the viral long-terminal repeats; (b) allowing the retrovirus to integrate into a gene of the animal or the ancestor thereof; and (c) determining the expression pattern of the nucleic acid sequence in the animal, wherein the

expression pattern of the nucleic acid sequence mimics the expression pattern of the gene. The animal may be, for example, a mouse, zebrafish, pufferfish, medaka, frog, fly (e.g., a fruit fly), goat, sheep, cow, pig, or chicken. The nucleic acid may include a reporter gene.

5 In a fifth aspect, the invention features a method for producing a transgenic non-human animal. this method includes (a) introducing into an ancestor of the animal a recombinant retrovirus that includes (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-terminal repeats; and (vi) nucleic acid sequence between the
10 splice acceptor and the splice donor, the nucleic acid sequence encoding a polypeptide in the direction opposite to the direction of transcription from the viral long-terminal repeats; and (b) allowing the retrovirus to integrate into the genome of the ancestor thereof. The animal may be, for example, a mouse, zebrafish, pufferfish, medaka, frog, fly (e.g., a fruit fly), goat, sheep, cow, pig,
15 or chicken.

 In a sixth aspect, the invention features a method for introducing a nucleic acid sequence into a cell. This method includes contacting the cell with a recombinant retrovirus including (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-
20 terminal repeats; and (vi) the nucleic acid sequence, and allowing the retrovirus to infect the cell. The cell may be *in vitro* or *in vivo*.

 In a seventh aspect, the invention features a method for identifying a high-titer virus producer cell line, including determining by quantitative PCR the ratio of viral DNA to a control DNA in the cell line. In a preferred
25 embodiment, the control DNA is a single copy gene.

 In an eighth aspect, the invention features a high-titer virus producer cell line identified by determining by quantitative PCR the ratio of viral DNA to a control DNA in the cell line.

In a ninth aspect, the invention features a virus produced by the cell line of the eighth aspect.

In a tenth aspect, the invention features a method for performing gene therapy on a mammal (e.g., a human), including administering the virus of the eighth aspect to the mammal.

In an eleventh aspect, the invention features a method for determining the level of recombinant retroviral infection in a sample from an animal, including determining by real-time quantitative PCR the ratio of viral DNA to a control DNA in the sample. The animal may be, for example, a human or a non-human (e.g., a mouse, zebrafish, pufferfish, medaka, frog, fly, goat, sheep, cow, pig, or chicken). In one preferred embodiment, the animal is a human who has undergone or is undergoing gene therapy, and the method is to monitor the efficacy of treatment.

By "branch point sequence" is meant a consensus nucleic acid sequence that is recognized in the circularization step in the mRNA splicing reaction (the step in which the 5' end of the intron forms a bond with an adenosine approximately 25 nucleotides upstream of the 3' splice site).

By "polypyrimidine tract" is meant a stretch of about 15 nucleotides that are all either cytosine or uracil (cytosine or thymidine in the DNA vector) and that lie between the branch point and the 3' splice site.

By "splice acceptor" is meant the nucleic acid sequence at the 3' splice site; this can be used to refer to merely the minimal requirements (an AG dinucleotide or an AG dinucleotide in the context of a short consensus sequence), or more generally to refer to a longer sequence encompassing 30-40 nucleotides upstream of this. A splice acceptor can also include a branch point sequence and a polypyrimidine tract.

By "splice donor" is meant the sequence of a 5' splice site; this may refer to either a short consensus sequence, generally beginning with AGGU

(AGGT in the DNA vector), or a longer sequence, including specific sequences upstream of the AGGU, that might increase the efficiency of the use of this splice site.

By "LTR" is meant long-terminal repeat of a retrovirus, a repeated sequence at both ends of the retrovirus that is required for many steps in the life cycle of the virus, including production of the viral RNA genome in the virus producing cells, proper reverse transcription to create the double stranded DNA provirus in the infected cell, and integration of the provirus into the infected cell's chromosome. The sequence of the 3' LTR in the producer virus-producing cell determines the sequence of both LTRs in the final integrated provirus. Thus the 5' LTR may be altered in some way to affect the production of the viral RNA genome in the producer cells without this alteration being evident in the infected cell; conversely, any alteration of the 3' LTR will be reflected at both LTRs in the infected cell.

By "reporter gene" is meant any gene which encodes a product whose expression is detectable and/or quantifiable by immunological, chemical, biochemical, biological, or mechanical assays. A reporter gene product may, for example, encode a protein having one of the following attributes, without restriction: fluorescence (e.g., *gfp*), enzymatic activity (e.g., *lacZ*/ β -galactosidase, luciferase, chloramphenicol acetyltransferase), toxicity (e.g., ricin), or an ability to be specifically bound by a second molecule (e.g., a FLAG epitope, a myc epitope, a HA epitope, biotin, or a detectably-labelled antibody). It is understood that any engineered variants of reporter genes, which are readily available to one skilled in the art, are also included, without restriction, in the foregoing definition.

By "splice enhancer" is meant a sequence that resides upstream of a splice donor, and that increases the efficiency of the use of that splice site.

By "exonic sequence" is meant sequences that remain in the mRNA following the splicing reaction (i.e., the sequences in the mature mRNA).

By "gene trap" is meant a vector containing a nucleic acid sequence gene that can only be expressed when it has inserted into a gene. For example, the nucleic acid sequence might lack a promoter, such that it must integrate downstream of the promoter of an endogenous gene in order to be expressed (often referred to more specifically as a "promoter trap"). Similarly, it may lack a promoter, but contain a splice donor, such that if it integrates into an intron of an expressed gene, it will be spliced into the mature RNA of that gene. Alternatively, a gene trap can contain a promoter, but lack other regulatory elements required for efficient gene expression (and thus detection of the reported gene). For example, it can contain a weak promoter which is only activated when integrated near an enhancer (often referred to as an "enhancer trap"), or contain a strong promoter but lack a polyadenylation signal, thus requiring integration upstream of a functional polyadenylation signal for proper expression.

By "quantitative PCR" is meant a use of the polymerase chain reaction under conditions which can reflect the amount of target sequence (the sequence which is recognized and amplified) in the starting material. This can be achieved, for example, in real-time with the TaqMan system using a Prism™ 7700 PCR machine (Perkin-Elmer/ABI Biosystems), or by running conventional PCR reactions at a limited number of cycles such that amplification is still in the linear range and analyzing the products of the reaction by either ethidium bromide staining after electrophoresis through agarose, or more sensitive means such as hybridization of radiolabeled or fluorescent probes.

By "recombinase" is meant an enzyme which catalyzes DNA recombination reactions, usually in a sequence-dependent manner. This

includes the site-specific recombinases from bacteriophage P1 (Cre) and yeast (FLP and HO).

By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By "high-titer virus producer cell line" is meant a cell line, isolated from a parent cell line, that has a viral DNA: single copy gene ratio that is among the top 50% of ratios from cell lines derived the parent cell line. Preferably, the cell line is in the top 33%, more preferably the cell line is in the top 25%; and most preferably the cell line is in the top 10%.

The invention features new gene trap cassettes for retroviral vectors. These cassettes are designed to increase splicing efficiency and to increase viral titer. As a result, the invention provides improved methods for introducing or mutating genes by retroviral-mediated gene trap techniques. The cassette and the methods for its use have a broad range of uses and compatible animal hosts.

The invention also provides an improved method for determining viral DNA content in a sample. The method is applicable for identifying high-titer virus producer cell lines (for the production of recombinant viruses for applications that are titer-dependent, such as gene therapy and other types of gene delivery), as well as for quantifying viral DNA in a sample for diagnostic purposes.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

Brief Description of the Drawings

Fig. 1 shows a schematic illustration of a recombinant retroviral vector containing the GT cassette. The viral LTRs are operably linked to lacZ

that has a nuclear localization signal (nIacZ). The direction of transcription from the LTRs is left to right. The GT cassette is in the opposite orientation. When the retrovirus integrates in the proper orientation into an active gene, the GT cassette is spliced into the transcript as a result of the splice acceptor (SA) and splice donor (SD) in the cassette.

Fig. 2A shows a schematic illustration of a GT cassette that contains branch point sequence (BPS); a polypyrimidine tract ((Py)_n); a splice acceptor (SA); exonic sequence; a splice enhancer; and a splice donor. The cassette is designed to result in truncation of the interrupted gene by the introduction of a stop codon or a frameshift in the transcript.

Fig. 2B shows a schematic illustration of a GT cassette that is identical to that shown in Fig. 2A, except that it also contains a reporter gene. The reporter gene can be followed by a stop signal. Alternatively, the translation can continue from the reporter sequence into the interrupted gene sequence. In the latter case, translation of the interrupted gene may produce a functional protein.

Fig. 2C shows a schematic illustration of a GT cassette that is identical to that shown in Fig. 2B, except that the reporter gene has been replaced by nucleic acid sequence encoding bacteriophage P1 Cre recombinase.

Fig. 3 shows a schematic illustration of quantitative real-time PCR. An oligonucleotide probe, nonextendable at the 3' end, labeled at the 5' end, and designed to hybridize within the target sequence, is introduced into the PCR assay. Annealing of the probe to one of the PCR product strands during the course of amplification generates a substrate suitable for exonuclease activity. During amplification, the 5'→3' exonuclease activity of a suitable DNA polymerase degrades the probe into smaller fragments that can be differentiated from undifferentiated probe. Measurement can be made, for example, on an ABI PRISM™ 7700 sequence detection system

Fig. 4 is a graph showing of the range of PCR titering. The results are linear over a broad range of virus concentration.

Fig. 5 shows a schematic illustration of the steps for isolation of high-titer virus producer cell lines.

5 Fig. 6 is a graph showing the correlation between PCR titer and recombinant retroviral infection in zebrafish embryos.

Detailed Description

We have discovered new retroviral vectors useful for gene-trapping in multicellular animals, and a method for identifying high-titer retroviral producer cell lines that may be used in any retroviral system.

10 Due to the difficulty in identifying mutated genes in zebrafish, the genetic basis of the vast majority of mutant phenotypes has not been determined. To facilitate the cloning of embryonic lethal mutations in the fish, we previously developed a method of insertional mutagenesis using mouse retroviral vectors pseudotyped with VSV-G envelope protein. The presence of
15 the VSV-G protein allows the retrovirus to enter zebrafish cells. When virus is injected into embryos at the blastula stage, cells that are destined to become germ line are among the cells that are infected. Every injected egg develops to be a founder fish that transmits proviral DNA to its progeny. On average, each
20 founder transmits about 12 proviral insertions. Most transgenic F1s inherit a single provirus, but some inherit two to four proviruses.

In an initial screen to determine if proviral insertions were mutagenic, we isolated seven insertional mutants. Six were embryonic lethal mutations, while one was a dominant, homozygous viable, adult mutation. The
25 frequency was approximately one insertional mutation per seventy proviral insertions. The seven mutants isolated to date fall into one of the three classes described above: one mutant displays apoptosis in the central nervous system,

three mutants display multiple defects, while three mutants have developmentally specific mutations.

Insertional mutagenesis allows for rapid cloning of the mutated genes. So far we have cloned candidate genes for six of the seven insertional mutants. Although, with present techniques, it is theoretically possible to generate enough insertions in the fish germ line to mutate all the genes, it is laborious to breed the fish harboring insertions to homozygosity in order to determine which have induced (i.e., recessive) embryonic mutations. In addition, the frequency of mutagenesis with the available viruses is rather low and integration events that occurred in large introns may not perturb gene expression and, thus, may not always appear mutagenic.

We have produced a series of new recombinant retroviruses that provide several advantages for use in mutagenesis experiments. Like previously-produced recombinant retroviruses, these new viruses have an intact $\psi+$ and MoMLV LTR sequences to maintain their full potential to be packaged at high titer in virus-producing cell lines and to be integrated in infected cells. They can also encode the β -galactosidase protein, starting at the ATG of the gag gene. In contrast to previously-produced recombinant retroviruses, however, viruses described herein also contain a novel gene-trapping module (named GT). The viruses that include this gene-trapping module differ from existing gene-trapping viruses in several regards. The general improvements are shown in Fig. 1 and are described below.

In contrast to other gene-trapping viruses, the viruses of the present invention have all of the elements necessary for efficient RNA splicing, including branch-point sequence, polypyrimidine tract, and a splice acceptor and splice donor flanking a mini-exon (see below). The presence of these elements facilitates recognition of and splicing with the inserted exon with the trapped gene.

Gene-trapping viruses normally have a terminal exon encoding a reporter gene followed by polyadenylation sequence, thus leading to a truncation of the mRNA and utilization of a polyadenylation sequence that is not endogenous to the gene that has been trapped. In contrast, the viruses
5 described herein have a mini-exon between the splice acceptor and splice donor. Transcription continues from an endogenous exon, through the mini-exon, and to the next exon of the trapped gene. Moreover, polyadenylation is via the endogenous polyadenylation sequence. The utilization of the endogenous polyadenylation sequence is likely to increase the amount of
10 mRNA that is produced.

The artificial mini-exon also harbors a pyrimidine-rich splice enhancer from avian sarcoma leukemia virus (ASLV) to augment its recognition by cellular RNA splicing machinery. In addition, the mini-exon encodes a small peptide epitope, such the FLAG epitope (DYKDDDDK) in one, two, or
15 all three reading frames. When the provirus integrates in an intron in the correct orientation, the artificial exon is spliced into the mRNA. If desired, the mini-exon can be designed so that it causes a frameshift mutation.

An additional advantage of the gene trap cassette described herein is its small size, which contributes to the virus having a high titer. The increased
20 titer, in turn, allows for more integration events per animal, which is important in the generation of mutants.

Cloning of the gene into which the virus integrates can be performed by RACE (e.g., 3' RACE or 5' RACE (Rapid Amplification of cDNA Ends)). The occurrence of an integration event can also be detected by RT-PCR, *in situ*
25 RNA hybridization, or immunodetection using antibodies against a peptide epitope.

We have demonstrated that the artificial exon is spliced into endogenous transcripts in the progeny of founder zebrafish injected with a

retrovirus containing the GT module. We mated male founder zebrafish with wild-type females, and isolated total RNA from a pool of 20 progeny harvested at 1-day or 5-days of age. After the RNA was digested with RNase-free DNase I to eliminate DNA contaminants in the samples, we then performed 5' RACE to determine if the mini-exon was present. We used an oligonucleotide that is specific to the mini-exon as a primer (primer 1) for cDNA synthesis. After tailing of the cDNA with dCTP, we PCR-amplified the cDNAs with a second mini-exon-specific primer (primer 2) internal (i.e., 5') to primer 1 and a 5' RACE abridged anchor primer. A nested PCR was then performed on the PCR products using a third mini-exon-specific primer (primer 3) and an abridged universal amplification primer. The PCR products were gel-purified and sequenced to determine whether the amplified products were of authentic hybrids of the mini-exon and mRNA. Of the 15 male founders analyzed, all but one yielded a discrete band in an agarose gel. Three of the 15 samples have fragments from two viral integration events, as determined by sequence analysis. Thus, there is, on average, at least one detectable gene-trap event per founder. Similar conclusions were reached by analyzing RNA from unfertilized eggs of female founders.

The viruses described above have many uses, including gene trap-mediated mutagenesis, gene expression analysis, and gene delivery. The viruses can encode a site-specific recombinase, allowing for conditional mutation of a gene that has sites recognized by one of the foregoing recombinases. Moreover, the viral vector can itself include loxP or FRT sites, such that the reporter gene or the entire viral vector can be removed from the host genome. Each of these uses is discussed in detail below.

Mutagenesis

The viral vectors containing the GT cassette are useful for gene trap-mediated insertional mutagenesis. Animals containing one or more proviral insertions are produced using standard techniques known to those skilled in the art (e.g., Couldrey et al., Dev. Dyn. 212:284-292, 1998). In one example, the GT cassette includes, between the splice acceptor and splice donor, nucleic acid sequence encoding a stop codon in one, two, or all three reading frames (Fig. 2A). Alternatively, the mini-exon itself can, by virtue of its nucleic acid length not being a multiple of three, lead to a frameshift that should result in a premature stop or non-functional protein. Depending on the insertion site, integration of the provirus may lead to a gene mutation and a resulting gross morphological or physiological phenotype.

Gene expression analysis

Determination of the expression pattern of the interrupted gene can also be performed using a viral vectors containing a GT cassette (Fig. 2B). This determination is possible whether the insertion of the provirus resulted in a mutation or not. In one example, the mini-exon contains nucleic acid sequence encoding a reporter polypeptide, such as a peptide epitope (e.g., FLAG, HA, or myc). It is preferred if all three reading frames encode a reporter polypeptide, as this assures the proper translation of a reporter polypeptide. Moreover, as we have found that smaller GT cassettes result in higher viral titer, small reporter polypeptides (such as the foregoing peptide epitopes) are preferred. Expression of the reporter polypeptide will be under control of the promoter of the interrupted gene. Thus, detection, such as immunodetection of a peptide epitope, of the expression pattern of the reporter polypeptide will reveal the expression pattern of the interrupted gene. Other methods of detection include, but are not limited to, RT-PCR, *in situ* hybridization, western blot analysis, and

detection of enzymatic activity or fluorescence.

Gene delivery

Another use of the GT cassette-containing retroviruses described herein is to express a nucleic acid of interest in a spatio-temporally dynamic manner. The disadvantages to this approach, in comparison to expressing a gene of interest from a defined promoter element, are that one cannot predict from the outset the resulting expression pattern, or if the interruption of a gene will result in a loss of function of that gene. The advantages include the relative ease of producing large numbers of lines of animals, and the ability to achieve an expression pattern not previously identified with a known promoter. The method is, in essence, identical to that described in the previous paragraph, except that the nucleic acid encoding the reporter polypeptide is replaced with the nucleic acid of interest. The expression pattern can be determined using standard techniques such as, for example, RT-PCR, *in situ* hybridization, immunohistochemistry, and western blot analysis.

Recombinase delivery

One particular set of preferred nucleic acids those encoding site-specific recombinases, such as the bacteriophage P1 Cre recombinase or FLP recombinase that have enzymatic activity when expressed in a vertebrate (Buchholz et al., Nuc. Acids Res. 24:4256-4262, 1996). Each of these recombinases recognizes a sequence motif, loxP and FRT, respectively; DNA flanked by two such sequence motifs is excised by the recombinase. Recently, strategies have been developed in which conditional expression of a gene is regulated by recombinase activity. For example, a mouse having loxP sites, engineered into introns flanking exon 2 of gene X, can be mated with a second mouse expressing Cre recombinase only in the central nervous system. In the progeny from this mating, cells in the central nervous system will have Cre

recombinase activity, and exon 2 of gene X will be excised. In all other tissues, exon 2 will not be excised.

Using the gene trap-mediated gene expression methods of the previous section, lines of animals can be produced, each of which having a unique expression pattern of recombinase (Fig. 2C). These animals can be mated to mice having loxP or FRT sites flanking a gene or gene segment, resulting in numerous lines of animals, each having a different pattern of gene inactivation.

Other uses of recombinase-mediated regulation of gene expression are described in Kilby et al., Trends Genet. 9:413-421, 1993. These uses are hereby incorporated by reference. Moreover, one skilled in the art will recognize that the use of other site-specific recombinases is also within the spirit of the invention.

LoxP or FRT sites

The viruses of the invention can also have loxP or FRT sites themselves (Russ et al., J. Virol. 70:4927-4932). For example, placement of loxP sites in the 5' and 3' LTRs allows for reversible gene interruption. In this example, the provirus has inserted into gene Y of a mouse, resulting in a truncation of the encoded protein and loss of function. The mutant mouse is then mated with a second mouse, in which Cre recombinase is expressed throughout the animal. In the progeny, Cre recombinase will excise nearly the entire proviral sequence, resulting in restoration of protein function.

The methods and viruses described herein have uses in a wide variety of animals that can be infected with retroviruses, including animals used in scientific research (e.g., mice, zebrafish, pufferfish, medaka, frog, and fly), and those with commercial value (e.g., goats, sheep, cows, pigs, and chickens). The GT cassette can be readily adapted to any retroviral vector, and, if

required, the VSV-G viral envelope can be employed for infection of nearly any vertebrate cell.

Method for producing high-titer virus producer cell lines.

In order to acquire enough viral particles for gene therapy, mutagenesis, or any other use of retroviral vectors, it is crucial to first obtain a virus producer cell line that yields very high titer.

A producer cell line is usually selected from a pool of candidates. Traditionally, it involves serial dilution of conditioned medium from individual candidate clones. These dilutions are then used to infect indicator cells (e.g. NIH 3T3), and the number of infection events is determined by counting the clones that express a viral marker (e.g., a visible marker such as lacZ or a selectable marker such as neo). This approach takes more than a week from time of infection to colony counting, and as a result requires intensive tissue culture manipulations to maintain the candidates in the duration. Consequently, usually only a few dozen candidate clones are usually tested. While this might result in a high-titer producing clone, it is statistically more likely that a larger initial pool of candidates will result in a better clone.

We have developed a high throughput screening method to isolate retrovirus producer cell lines. Using this method, a high-titer GT virus producer cell line was quickly selected from 230 candidates. This method uses real-time quantitative PCR analysis to compare the number of proviral insertions in target cells transduced by the conditioned medium of individual candidate. The PCR assay uses ABI PRISM™ 7700 sequence detection system (PE Applied Biosystems, Foster City, CA) to quantify PCR product accumulation through a dual-labeled virus specific fluorogenic probe (i.e., TaqMan Probe). Briefly, an oligonucleotide probe, nonextendable at the 3' end, labeled at the 5' end, and designed to hybridize within the target sequence,

is introduced into the PCR assay. Annealing of the probe to one of the PCR product strands during the course of amplification generates a substrate suitable for exonuclease activity. During amplification, the 5'→3' exonuclease activity of a DNA polymerase (e.g., Taq polymerase) degrades the probe into smaller fragments that can be differentiated from undifferentiated probe (Fig. 3) (Holland et al., Proc. Acad. Natl. Sci. USA 88:7276-7280, 1991; Heid et al., Genome Res. 6:986-994, 1996). Quantitative data are derived from a determination of the cycle at which the fluorescence reaches a preset detection threshold. The earlier the threshold is reached, the more target in the sample. The assay is very accurate and reproducible (Fig. 4). The selection procedure is outlined below.

Candidate clones were made by infecting 293bsr, a packaging cell line, with GT virus followed by fluorescent activated cell sorting (FACS) of infected cells. GT virus was prepared by transient co-transfection of pCMV-GT and pCMV-VSV-G. FACS was performed two days post-infection after loading infected cells with FDG, a membrane-permeable fluorogenic substrate of β -galactosidase. Cells with the strongest fluorescence (top 0.5%) were collected and seeded individually into 96-well plates (Fig. 5). Once grown to confluence, cells in each well were transferred to a compartment in 24-well plates. When confluent again, cells were divided into six parts and each part seeded in a well of one of six 96-well plates. The remaining cells were maintained in the same compartment. The next day, the cells were transfected with pCMV-VSV-G (by calcium phosphate co-precipitation or lipofectamine), then cultured in 100 μ L of medium.

One important aspect of the method described herein is, although VSV-G coat viruses can infect a broad range of hosts, the *in vitro* titering of virus should be performed on cells of the same animal type will be infected *in vivo*. Hence, two days after transfection, 25 μ L of the medium in each well was

taken to infect PAC2 cells (a zebrafish cell line) in 96 well plates. Two days post-infection, the PAC2 cells were lysed in 25 μ L GNT buffer. Lysate (2 μ L) from each well was analyzed by simultaneous real-time quantitative PCR analysis for viral DNA and an endogenous single-copy gene, RAG1. The ratio of viral DNA to RAG1 DNA were then determined, and the top 10% in each
5 transfection method of the first 230 clones were kept for second round of selection. After repeating the above procedure, the top six clones from each transfection method were chosen for large-scale preparation.

The titer of each of the clones was determined by quantitative PCR,
10 and quantitative Southern analysis in injected fish larvae. As predicted, there is a very strong correlation between the two assays (Fig. 6). The top clone in each category, clone 186 for lipofection and clone 202 for calcium phosphate, both yield virus with titers comparable to existing virus without any optimization. Because most of the manipulations are carried out in 96-well
15 microtiter plates, and one plate can be analysed in less than 2 hours, a large number of clones can be selected in a very short period of time. In addition, no extra cell culture maintenance is needed as it only takes 5 to 6 days to identify the top clones after splitting the cells for transfection. By then, the left-over cells have just become confluent.

20 Clone 186 was selected from 230 candidates to make GT virus by lipofection. Because lipofection is a more robust method of transfection, it was chosen for subsequent studies. To optimize lipofection conditions, we compared the titer of virus made by the combination of 5, 10, 15, and 20 μ g of pCMV-VSV-G DNA per 15 cm plate at ratios of DNA to lipofectamine of
25 1:10, 1:14, 1:20, and 1:30 (w/w). Of those, 5 μ g DNA, 75 μ l lipofectamine per plate produced virus with high titer and low toxicity. At this condition, clone 186 produced viruses that generate, in injected fish, an average of 20 proviral insertions per cell after one round of injection, compared to 4.5 after two

rounds of injection using the existing control cell line. In addition, the survival rate following injection also increased from 40% to more than 65%. The combination of these factors has increased the weekly output of adult founders from 600 to more than 3000.

5 This rapid and high throughput assay for high-producer cells has other advantages over conventional screening procedure. First, it does not require a marker gene in the virus. Most retroviral vectors, in particular those for gene therapy purposes, contain a selectable marker gene specifically for selecting a producer cell line. Inclusion of the extra sequence may not only
10 lower the titer by promoter suppression, but also increase the immunogenicity of host cells. It is desirable to make viral vectors that do not have a selectable marker. The method described herein is useful for selecting good producer cells for these very viruses. Second, the method selects clones based on the titer on cells derived from the target species. Third, the method described
15 herein is more reliable than methods that measure viral RNA content in conditioned medium. Viral RNA is a poor predictor of infection, especially for cells from non-mammals such as zebrafish. Finally, it is much less labor-intensive than methods that uses competitive quantitative PCR analysis.

 The reliable production of very high-titer stocks of retroviral vectors
20 is essential for several important applications, including human gene therapy and the production of transgenic animals. Transgenic animals, in turn, have two important uses with commercial implications: the expression of novel genes in an animal and insertional mutagenesis.

Human gene therapy

25 Production of sufficient amounts of high-titer virus to infect large numbers of specific cell populations for the treatment of human diseases has been a challenging problem. Clearly, the higher the viral titer for the particular

cell or tissue to be infected, the better the gene delivery. The technology described here allows one to rapidly identify cell clones, harboring the viral vector of choice, that produce high titers of virus that are capable of infecting the cell of interest.

5 Production of transgenic animals

Insertional mutagenesis makes it possible to rapidly clone genes required for any biological process of interest. To isolate mutations in the genes of interest requires that one generate a large number of proviral insertions in the germ line of the species of interest. Animals harboring the insertions are then bred so as to bring the insertions to homozygosity and mutant animals are then identified by screening for the phenotype(s) of interest. The method described herein allows one to isolate clones of cells producing very high-titer stocks of virus for infecting virtually any animal species. Thus the technology opens the possibility of using retroviral vectors to perform insertional mutagenesis in a wide variety of animal species.

As described above, the insertion of large nucleic acid sequences decreases the viral titer. The cell lines identified by the method described herein provide a means for producing high-titer recombinant virus containing large inserts. This allows for gene mis-expression, production of transgenic animals, or gene therapy in cases where the viral titer produced using previous means was inadequate.

The method for quantifying virus described herein has other uses. Cells, such as stem cells, that might be infected before being transplanted into an animal, can be quickly assayed for viral DNA content. Similarly, diagnostic assays for infection levels (either due to viral infection (e.g., HIV) or as a means to follow the progress of gene therapy) can also use this technique. In another example, quantifying recombinant retroviral DNA is useful in

determining whether a batch of injected animals are likely to have a high frequency of proviral insertions. In generating insertional mutants, we have observed a strong correlation among animals injected with a given viral preparation by a given person. Accordingly, by assessing the viral DNA in a small subset of the injected animals, we can readily and quickly determine whether it is worth the time, money, and effort to maintain the remaining animals. While this is important for fish, it is even more important for higher animals (e.g., mice) that have longer development times and increased maintenance costs.

10 The particular cell clone described here, which produces very high-titer virus stocks for infecting the zebrafish germ line, contains a viral vector that includes a gene trap cassette. This cassette will likely lead to a large increase in the mutagenic frequency of the virus and will accelerate the identification of mutants and the subsequent cloning of mutated genes.

15 The expression of genes in transgenic animals can be valuable for basic research purposes and in some cases for commercial purposes directly. An example of the latter is the introduction of a gene coding for growth hormone into commercially valuable fish species to speed their growth. Vectors would be expected to express genes in a wide variety of fish as well as in other animals, such as chickens, cows, pigs, and sheep. The methods described herein are useful for generating cell clones producing high titers of any retroviral vector construct.

Other Embodiments

25 All patent applications and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent application and publication was specifically and individually indicated to be incorporated by reference.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover any variations, uses, or adaptations following, in general, the principles of the invention and including
5 such departures from the present disclosure within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth.

Other embodiments are within the claims.

What is claimed is:

1. A recombinant retrovirus comprising:
 - (a) branch-point sequence;
 - (b) a polypyrimidine tract;
 - (c) a splice acceptor;
 - 5 (d) a splice donor; and
 - (e) long-terminal repeats
2. The recombinant retrovirus of claim 1, wherein said splice acceptor and said splice donor flank nucleic acid sequence encoding a stop codon that is in frame with said splice acceptor.
- 10 3. The recombinant retrovirus of claim 1, further comprising a reporter gene.
4. The recombinant retrovirus of claim 3, wherein said reporter gene is in the direction opposite to the direction of transcription from the viral long-terminal repeats.
- 15 5. The recombinant retrovirus of claim 3, wherein said reporter gene is selected from the group consisting of *gfp*, *lacZ*, and a nucleic acid encoding myc epitope, a FLAG epitope, or a HA epitope.
6. The recombinant retrovirus of claim 3, wherein reporter genes are
20 in all three reading frames.
7. The recombinant retrovirus of claim 1, further comprising a splice enhancer.

8. The recombinant retrovirus of claim 7, wherein said splice enhancer is from the avian sarcoma leukosis virus.

9. The recombinant retrovirus of claim 1, further comprising exonic sequence between said splice acceptor and said splice donor.

5 10. The recombinant retrovirus of claim 1, further comprising nucleic acid sequence encoding a polypeptide encoded in the direction opposite to the direction of transcription from the viral long-terminal repeats.

10 11. The recombinant retrovirus of claim 10, wherein said polypeptide comprises a polypeptide selected from the group consisting of GFP, β -galactosidase, a myc epitope, a FLAG epitope, and a HA epitope.

12. The recombinant retrovirus of claim 10, wherein said polypeptide comprises Cre recombinase or FLP recombinase.

13. A method for performing gene-trapping in a cell, comprising:
15 (a) contacting said cell with a recombinant retrovirus comprising (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-terminal repeats; and (vi) a reporter gene in an orientation opposite to the direction of transcription from the viral long-terminal repeats, wherein said reporter gene is expressed if there is a gene-trapping event; and

20 (b) allowing said retrovirus to integrate into the genome of said cell.

14. The method of claim 13, wherein said cell is *in vitro*.

15. The method of claim 13, wherein said cell is *in vivo*.

16. A method for introducing a mutation into a gene in a cell,
comprising:

(a) contacting said cell with a recombinant retrovirus comprising: (i)
5 branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a
splice donor; and (v) viral long-terminal repeats, wherein said splice acceptor
and said splice donor flank nucleic acid sequence encoding a stop codon that is
in frame with said splice acceptor; and

(b) allowing said retrovirus to integrate into a gene of said cell,
10 wherein integration into said gene introduces a mutation into said gene.

17. The method of claim 16, further comprising (c) determining the
site of integration of said retrovirus.

18. The method of claim 16, wherein said cell is *in vitro*.

19. The method of claim 16, wherein said cell is *in vivo*.

20. A method for determining the expression pattern of a gene in a non-human animal, comprising:

(a) introducing into said animal or an ancestor thereof a recombinant retrovirus comprising (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-terminal repeats; and (vi) nucleic acid sequence between said splice acceptor and said splice donor, said nucleic acid sequence encoding a polypeptide in the direction opposite to the direction of transcription from said viral long-terminal repeats;

(b) allowing said retrovirus to integrate into a gene of said animal or said ancestor thereof; and

(c) determining the expression pattern of said nucleic acid sequence in said animal, wherein the expression pattern of said nucleic acid sequence mimics the expression pattern of said gene.

21. The method of claim 20, wherein said animal is selected from the group consisting of mice, zebrafish, pufferfish, medaka, frogs, flies, goats, sheep, cows, pigs, and chickens.

22. The method of claim 20, wherein said nucleic acid comprises a reporter gene.

23. A method for producing a transgenic non-human animal, comprising:

- (a) introducing into an ancestor of said animal a recombinant retrovirus comprising (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-terminal repeats; and (vi) nucleic acid sequence between said splice acceptor and said splice donor, said nucleic acid sequence encoding a polypeptide in the direction opposite to the direction of transcription from said viral long-terminal repeats; and
- (b) allowing said retrovirus to integrate into the genome of said ancestor thereof.

24. The method of claim 23, wherein said animal is selected from the group consisting of mice, zebrafish, pufferfish, medaka, frogs, flies, goats, sheep, cows, pigs, and chickens.

25. A method for introducing a nucleic acid sequence into a cell, said method comprising contacting said cell with a recombinant retrovirus comprising (i) branch-point sequence; (ii) a polypyrimidine tract; (iii) a splice acceptor; (iv) a splice donor; (v) viral long-terminal repeats; and (vi) said nucleic acid sequence, and allowing said retrovirus to infect said cell.

26. The method of claim 25, wherein said cell is *in vitro*.

27. The method of claim 25, wherein said cell is *in vivo*.

28. A method for identifying a high-titer virus producer cell line, comprising determining by quantitative PCR the ratio of viral DNA to a control DNA in said cell line.

29. The method of claim 28, wherein said control DNA is a single copy gene.

30. A high-titer virus producer cell line identified by determining by quantitative PCR the ratio of viral DNA to a control DNA in said cell line.

5 31. A virus produced by the cell line of claim 30.

32. A method for performing gene therapy on a mammal, comprising administering the virus of claim 31 to said mammal.

33. The method of claim 32, wherein said mammal is a human.

10 34. A method for determining the level of infection in an animal, comprising determining by real-time quantitative PCR the ratio of viral DNA to a control DNA in a sample from said animal.

35. The method of claim 34, wherein said animal is selected from the group consisting of mice, zebrafish, pufferfish, medaka, frogs, flies, goats, sheep, cows, pigs, and chickens.

1/5



Fig. 1

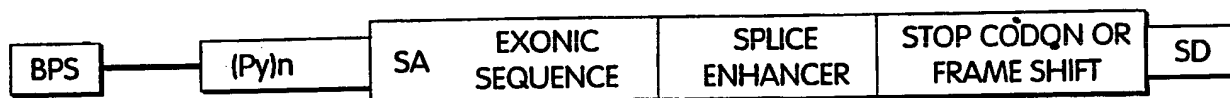


Fig. 2A

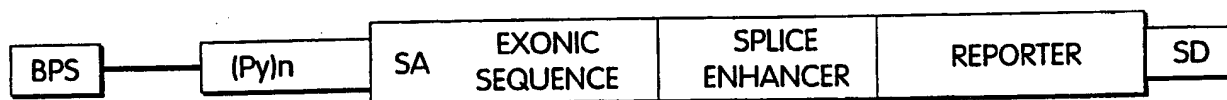


Fig. 2B



Fig. 2C

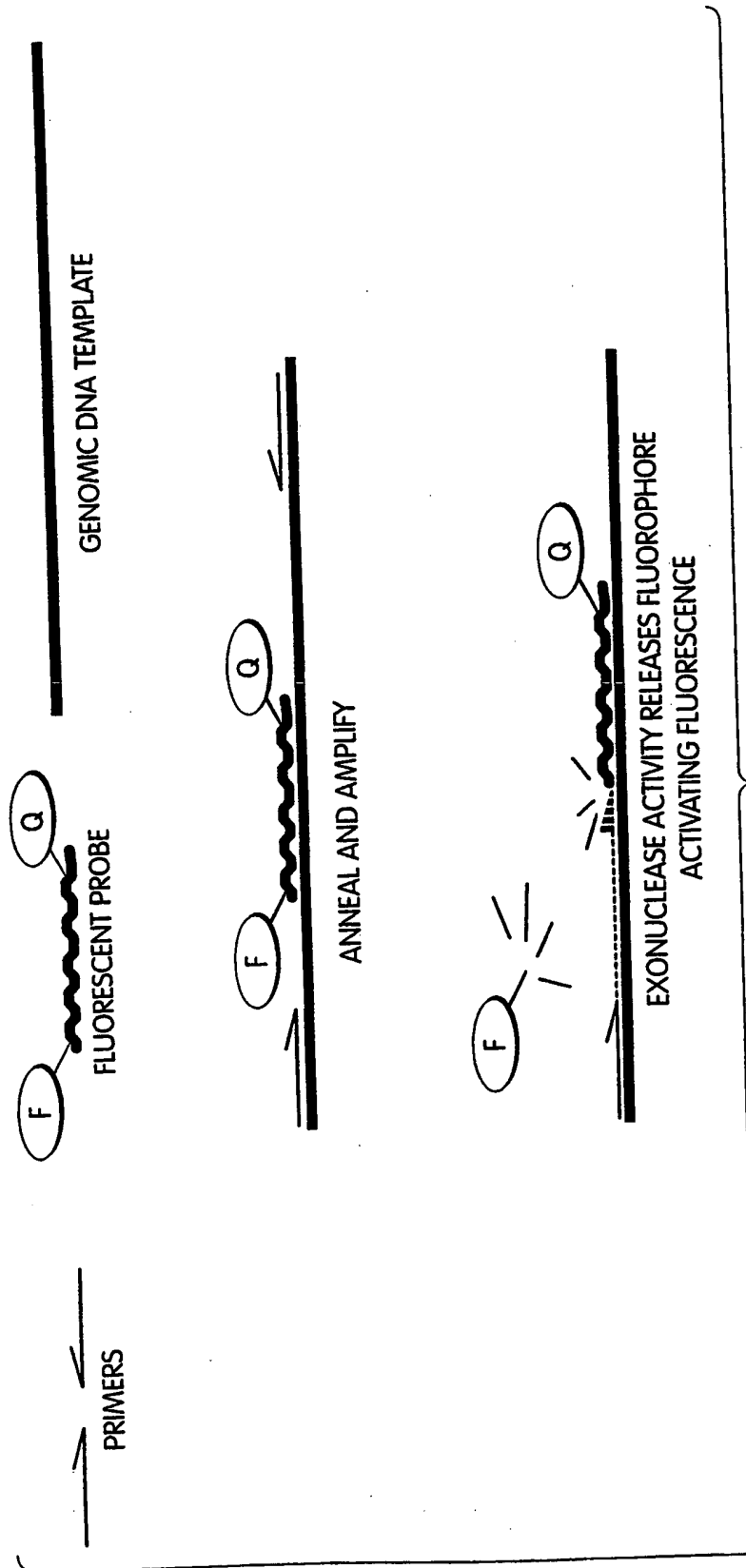


Fig. 3

3/5

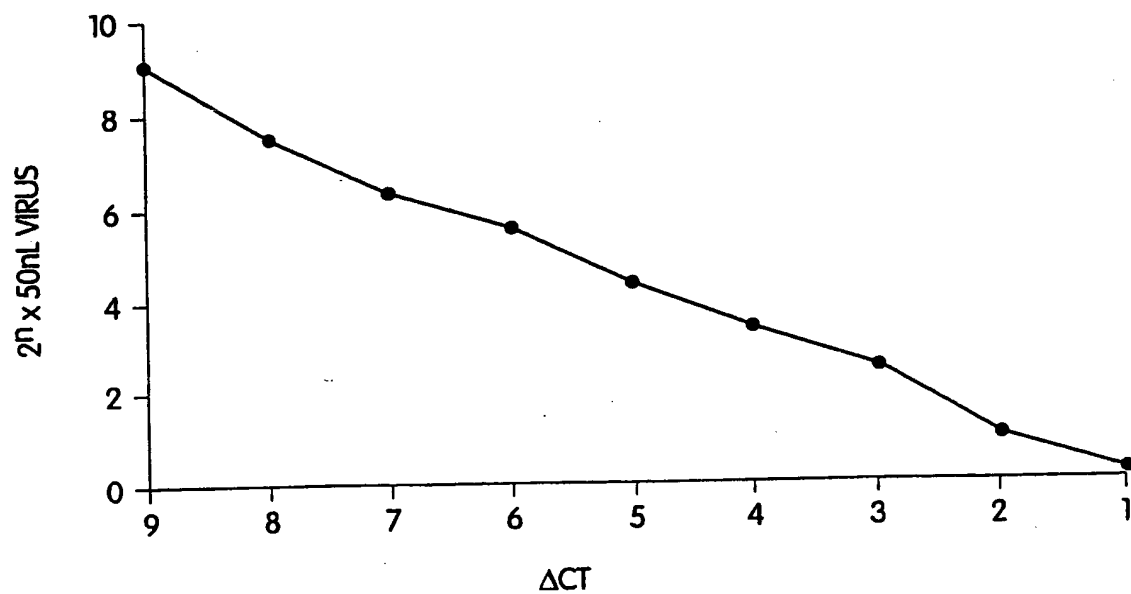


Fig. 4

4/5

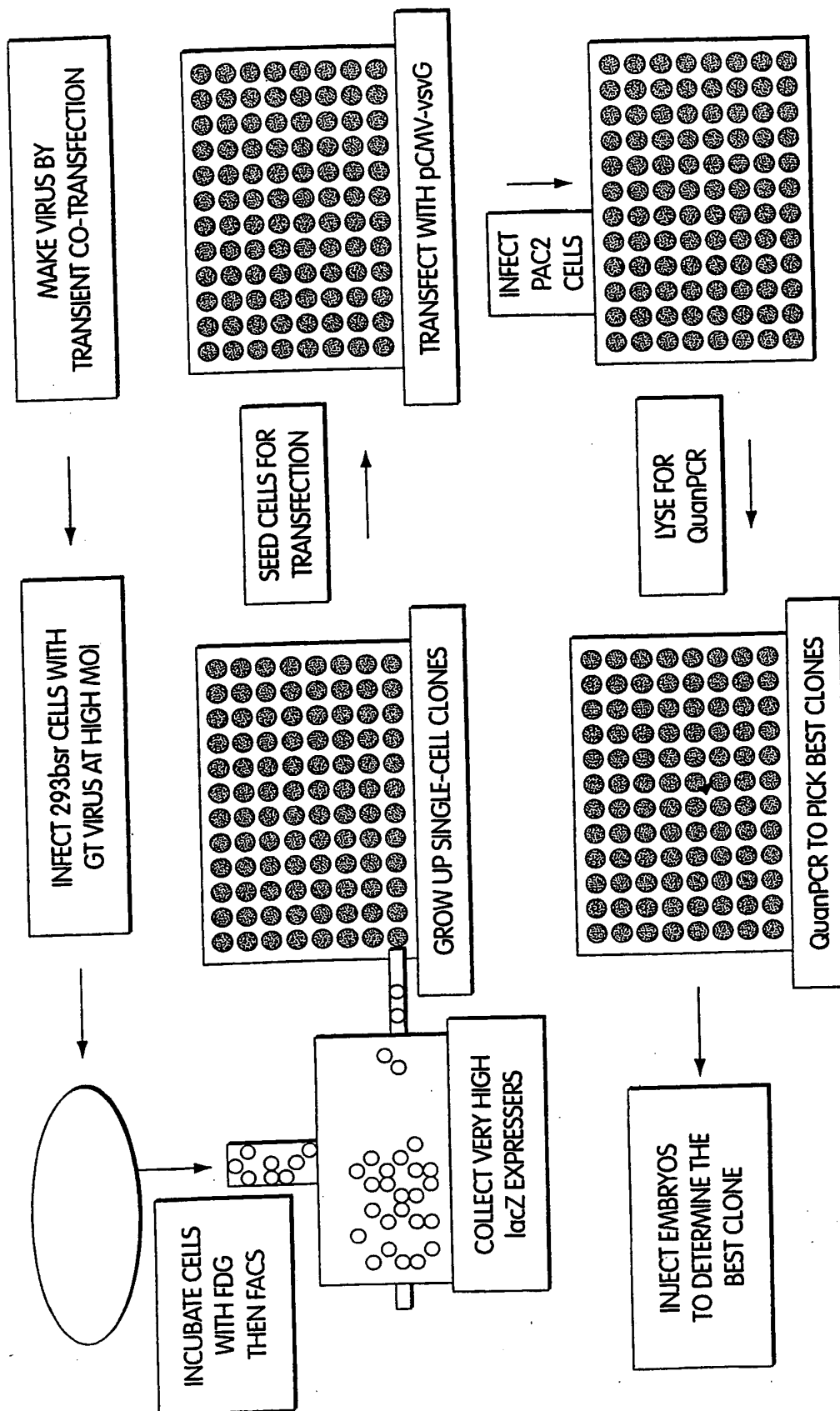


Fig. 5

5/5

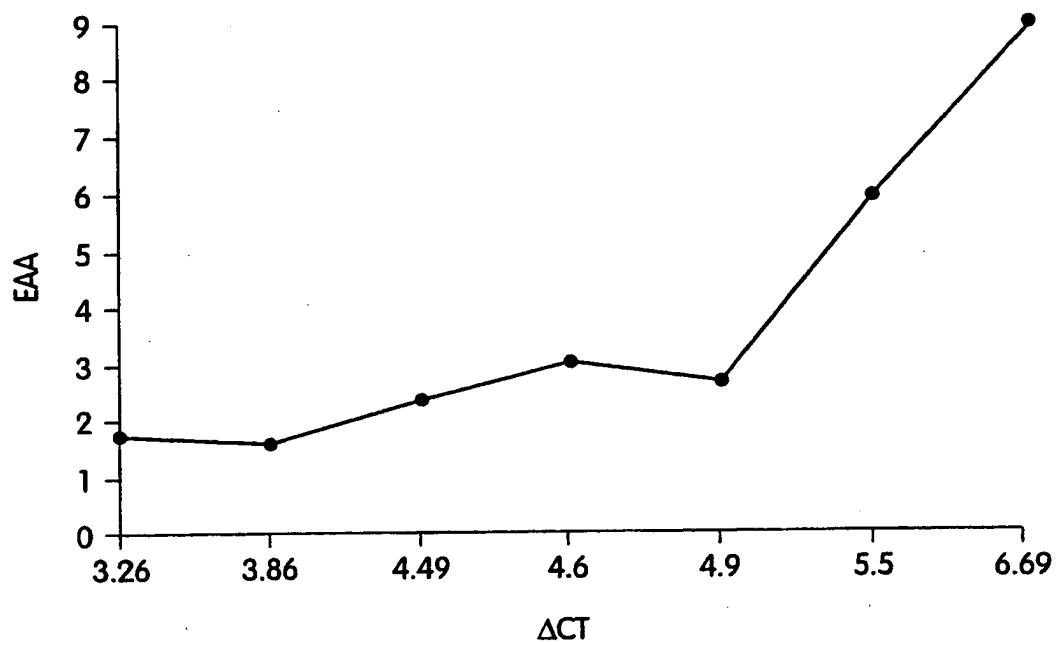


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/07841

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C 12N 15/00, 15/86; AOIN 63/00
US CL : 435/320.1, 456, 530; 424/93.2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1, 456, 530; 424/93.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Medline, caplus, biosis, embase, scisearch, USPAT

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BOUCK, J. Role of the Constitutive Splicing Factors U2AF65 and SAP49 in Suboptimal RNA Splicing of Novel Retroviral Mutants. The Journal of Biological Chemistry. 12 June 1998, Vol. 273. No. 24, pages 15169-15176, especially Abstract, Column 2 line 6.	1,8,15-19
Y	US 5,679, 523 A (LI et al) 21 October 1997, Col. 1 lines 38-43, lines 57-59; col 4 lines 5-12, lines 24-27; col 5 lines 55-60; col 8 lines 31-40, lines 63-67; col 13 lines 3-5; col 17 lines 3-5;	2-7, 9-14
X	BOUCK, J. et al. In vivo selection for intronic splicing signals from a randomized pool. Nucleic Acids Research. 1998, Vol. 26, No. 9, pp 4516-4523 especially Abstract.	1,7-8, 15-19

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*G* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
17 MAY 2000

Date of mailing of the international search report
11 JUL 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
Eleanor Sorbello
ELEANOR SORBELLO
Telephone No. (703) 308-0196